



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

What is the uncertainty in degree day projections due to different 2 calibration methodologies?

Citation for published version:

Holmes, C, Tett, S & Butler, A 2017, 'What is the uncertainty in degree day projections due to different 2 calibration methodologies?', *Journal of Climate*, vol. 30, no. 22, pp. 9059-9075. <https://doi.org/10.1175/JCLI-D-16-0826.1>

Digital Object Identifier (DOI):

[10.1175/JCLI-D-16-0826.1](https://doi.org/10.1175/JCLI-D-16-0826.1)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Climate

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



What is the Uncertainty in Degree-Day Projections due to Different Calibration Methodologies?

CAROLINE HOLMES AND SIMON TETT

School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom

ADAM BUTLER

Biomathematics and Statistics Scotland, Edinburgh, United Kingdom

(Manuscript received 18 November 2016, in final form 14 July 2017)

ABSTRACT


Degree-days are a temperature index used for understanding the impact of climate change. Different methods to deal with climate model biases, termed bias correction or more generally calibration, yield different projections of such indices, something not widely understood for temperature indices in many impact sectors. An analytical expression is derived for the expected value of degree-days given parameters of the underlying statistical distribution (assumed to be Gaussian). It is demonstrated that the uncertainty introduced by calibration methodology is driven by the magnitude of the nonlinearity in this expression. In a climate where mean temperature is, and remains, far from (approximately three standard deviations) the threshold used in defining the index, the equation is approximately linear, and methodological choice makes little difference relative to the absolute number of degree-days. However, case studies for U.K. cities London and Glasgow for heating and cooling degree-days (HDD and CDD; these are degree-day indices used in the estimation of energy use for heating and cooling buildings) demonstrate that, when temperatures are close to the threshold, unrealistic results may arise if appropriate calibration is not performed. Seasonally varying temperature biases in the 11-member perturbed parameter ensemble HadRM3 are discussed, and different calibration strategies are applied to this ensemble. For projections of U.K. HDD, the difference between results from simple and advanced methodologies is relatively small, as the expression for HDD is approximately linear in many months and locations. For U.K. CDD, an inappropriate method has a large relative impact on projections because of the proximity to the threshold. In both cases, the uncertainty caused by methodology is comparable to that caused by ensemble spread.

1. Introduction

It is well known that the demand for energy is related to temperature because of the use of energy for heating and cooling buildings (Taylor and Buizza 2003; Isaac and Van Vuuren 2009). Therefore, changes in future energy demand will be affected not only by socioeconomic factors such as economic and population growth but also by changes in climate. This has important implications for building design and planning. Moreover, this is an important example of a situation in which mitigation of and adaptation to climate change interact;

in a warming climate, society would need to adapt its energy use to cope with changed temperature. All other things being equal, reductions in energy use because of reduced need for heating would reduce emissions, thus contributing to mitigation. On the other hand, increases in energy use because of the need for cooling would increase emissions, making mitigation harder to achieve.

Indices based on meteorological variables are important tools for understanding societal impacts of climate change (e.g., Zubler et al. 2014; Harding et al. 2015), providing simple impact-relevant summaries of meteorological information. In general, degree-days are a measure of the sum of temperature deviations above or below a threshold (a more precise definition is given in section 2a). Heating

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Caroline Holmes, calmes@bas.ac.uk



degree-days (HDD) and cooling degree-days (CDD) are common temperature indices used to estimate the requirement of energy for space heating and cooling in order to maintain comfortable indoor temperatures (e.g., [Li et al. 2012](#)). Energy use for heating and cooling is calculated as a product of the degree-day (DD) index and factors describing energy efficiency, area to be heated per population member, and population size ([Isaac and Van Vuuren 2009](#)).

Climatologies and observed trends in HDD and CDD have been published for Scotland ([Sniffer 2014](#)) and the whole United Kingdom ([Jenkins et al. 2008](#)). Both HDD and CDD vary across the United Kingdom in a way that is consistent with geographical and altitudinal temperature variations. In the period 1961–90, mean annual HDD varied from below 2000 in London and the Channel Islands to over 4200 in high-altitude areas of Scotland. CDD are historically low, with climatological values for the same period of below 40 DD across the United Kingdom and less than 5 DD over much of Scotland, Northern Ireland, and north England ([Jenkins et al. 2008](#)). A more recent report ([Kendon et al. 2016](#)) demonstrates that, consistent with temperature increases, the later period 1981–2010 had less HDD across the region, with a U.K.-average reduction of 165 HDD between the periods. CDD increased in the same period in England and Wales, but in Scotland, where CDD remain rare, increases were not apparent. Both CDD and HDD display sizeable interannual variability ([Kendon et al. 2016](#)).

Projected future changes in HDD and CDD have been evaluated globally ([Isaac and Van Vuuren 2009](#); [Li et al. 2012](#)) and for specific locations in Europe ([Christenson et al. 2006](#); [Zubler et al. 2014](#); [Lemonsu et al. 2013](#)). [Day et al. \(2009\)](#) projected near-term changes in CDD in London, while [Chow and Levermore \(2010\)](#) incorporated DD-type indices into evaluation of near-term changes in heating and cooling demand for London (southeast England), Manchester (northwest England), and Edinburgh (southeast Scotland). However, there are no recent published studies giving projections of HDD and CDD for the United Kingdom as a whole. Heating demand is projected to decrease, with reductions of approximately 30% for both Paris ([Lemonsu et al. 2013](#)) and lowland Switzerland ([Zubler et al. 2014](#)) by the end of the twenty-first century, but cooling demand is projected to increase. In temperate regions such as the United Kingdom, the decrease in HDD dominates so net DD decrease ([Lemonsu et al. 2013](#); [Zubler et al. 2014](#); [Isaac and Van Vuuren 2009](#); [Li et al. 2012](#)). However, this may not be a relevant metric for understanding impacts on total energy demand because cooling demand is primarily met by electricity, but this is not the case for heating demand.

Climate model projections used to investigate future climate must be adjusted to allow for known deficiencies

(biases) of that model, a process referred to as calibration. Different calibration strategies arise from different assumptions about the future behavior of the real world relative to that of a climate model. These can produce different results ([Ho et al. 2012](#)). While there are considerable challenges inherent in producing bias-corrected projections of future climate, it is important to understand the possible consequences of different calibration methodologies.

For projections of DD specifically, various studies have noted that the standard method of calculating projected changes using climate models does not take account of climate model biases (e.g., [Zubler et al. 2014](#); [Erhardt 2015](#)) but without attempting to address this. While calculating the change in an index such as DD may be the most appropriate approach in some cases ([Hanlon et al. 2014](#)), for certain indices such approaches may produce very unrealistic results ([Hawkins et al. 2013](#)). For example, for the simple index “days above a temperature threshold,” consider a climate model that has a small cold bias relative to observations but assume that present climate never exceeds the threshold in the model or observations. [This example follows the discussion by [Hawkins \(2015\)](#).] The index is therefore zero in both model and observations; the bias in the index is zero. Let us assume that the temperature change projected by the model is the same as the (unknown) change in the real world. Consider then a case in which the model projects a small warming in the future, which is insufficient to give any days over the threshold in the model but would do so in the real world. The change in index calculated in the model is therefore demonstrably inaccurate given the underlying assumptions. It is furthermore reasonable to argue that the model is more likely to correctly represent change in an underlying temperature distribution (since it is a physically formulated model) than in an abstract index of that temperature distribution. This issue has been raised for threshold exceedances in the context of agriculture ([Hawkins et al. 2013](#)), but the importance of such a distinction for HDD and CDD has not been addressed previously in the literature.

The primary goal of this paper is to assess the effect of climate model temperature bias on HDD and CDD and on calibration methods. This is then used to gain understanding of this source of uncertainty in projections of future HDD and CDD in the United Kingdom. The paper is structured as follows: [Section 2](#) discusses different calibration methods, derives the analytical model for expected HDD and CDD and explores its properties, and describes the observational and climate model data used. [Section 3a](#) demonstrates the biases in mean temperature and subseasonal variance in the climate model. [Section 3b](#) explores the implications for future projections by

applying the analytical model, given sample biases in mean and standard deviation. Finally, projections of future U.K. HDD and CDD are presented using different calibration strategies applied to model data in [section 3c](#). Given the simplicity of the analytical model, implications of this work for other indices and regions are easy to deduce and are discussed in the discussion section.

2. Data and methods

This section begins by defining the indices of particular relevance to this paper, HDD and CDD. It then discusses four different calibration strategies that may be used to estimate the values of temperature indices in a future climate and their underlying assumptions. An analytical model for the expected value of HDD and CDD, given properties of the underlying temperature distribution, is derived; its properties, and how they may inform us as to the suitability of the simplest calibration strategies, are discussed. Finally, details of the observational and model data and regions of interest that will be used to apply the analytical model and calibration strategies are given.

a. Indices: HDD and CDD

There are a variety of definitions of HDD and CDD, although all are measures of the sum of temperature deviations above (for CDD) or below (HDD) a threshold. Consistent with the [Sniffer \(2014\)](#) analysis of historical Scottish degree-days, this study uses daily mean temperatures and thresholds $T_H = 15.5^\circ\text{C}$ for HDD and $T_C = 22.0^\circ\text{C}$ for CDD. For daily temperatures x_i , for days $i = 1, 2, \dots, N$,

$$\text{HDD} = \sum_{i=1}^N (T_H - x_i) H(T_H - x_i), \quad (1)$$

$$\text{CDD} = \sum_{i=1}^N (x_i - T_C) H(x_i - T_C), \quad (2)$$

where $H(x)$ is the Heaviside step function such that $H(x) = 0$ if $x < 0$ and $H(x) = 1$ otherwise. Therefore, CDD only occur on days above the threshold, and HDD only occur on days below the threshold. Many previous studies in Europe and the United States use different thresholds, often $T_C = T_H = 18.3^\circ\text{C}$ (65°F ; [Christenson et al. 2006](#); [Isaac and Van Vuuren 2009](#); [Zubler et al. 2014](#)).

b. Calibration strategies

In this context, a calibration strategy is a method that takes the future climate of a climate model, which is known to be biased in the historical period, and attempts to derive information about the true, unknown, future climate state. We consider the case in which the quantity for which projections are required is a general

temperature index that can be described as some function $g(\mathbf{x})$, where \mathbf{x} is a time series of observations.

Using standard statistical notation such that X denotes a random variable and \mathbf{x} denotes a sample from the distribution of X , we define

- \mathbf{x}_{oh} observed (“o”) temperatures for the historical (“h”) period;
- \mathbf{x}_{mh} modeled (“m”) temperatures for the historical period;
- \mathbf{x}_{mf} modeled temperatures for the future (“f”) period; and
- X_{of} unknown future observed temperature distribution, which we wish to estimate.

In this study, four calibration strategies are considered. Two are applied to the index only; the other two are applied to the underlying temperatures following the methods discussed in [Ho et al. \(2012\)](#). They are as follows:

- 1) Additive index correction: The historical bias in the index is removed from the modeled future value. The projected value of the index I is denoted $I_{\Delta I}$, where $I_{\Delta I} = g(\mathbf{x}_{\text{oh}}) + g(\mathbf{x}_{\text{mf}}) - g(\mathbf{x}_{\text{mh}})$.
- 2) Proportional index correction: The modeled future value is corrected by the historical proportional bias. The projected value of the index is denoted $I_{\times I}$, where $I_{\times I} = g(\mathbf{x}_{\text{oh}})g(\mathbf{x}_{\text{mf}})/g(\mathbf{x}_{\text{mh}})$.
- 3) Underlying change: The underlying temperatures are calibrated, assuming that the change from historical to future climate is not dependent on that historical climate (change independent of bias). Therefore, a mapping from historical to future climate, obtained from the model, can be applied to observed historical values. We assume that the change in mean and variance summarizes the change in the full distribution. Formally, using the subscripts defined above, a time series of future temperatures \mathbf{x}_{of} is obtained by the mapping $\mathbf{x}_{\text{of}} = \mu_{\text{mf}} + (\sigma_{\text{mf}}/\sigma_{\text{mh}})(\mathbf{x}_{\text{oh}} - \mu_{\text{mh}})$ ([Ho et al. 2012](#)). The index is then calculated as $g(\mathbf{x}_{\text{of}})$ and denoted $I_{\Delta T}$.
- 4) Underlying bias correction: The underlying temperatures are again calibrated. This time a subtly different assumption is made that the bias between observations and models is constant, that is, independent of time. In this case, a mapping from modeled to observed climate can be derived from the historical period and applied to future modeled climate. The mapping is $\mathbf{x}_{\text{of}} = \mu_{\text{oh}} + (\sigma_{\text{oh}}/\sigma_{\text{mh}})(\mathbf{x}_{\text{mf}} - \mu_{\text{mh}})$, and the index $g(\mathbf{x}_{\text{of}})$ is denoted I_{eT} .

Figure 1 of [Ho et al. \(2012\)](#) gives an illustration of the third and fourth methods. These two final methods differ because including the variance as well as mean makes the problem a nonlinear one; in the linear case (distributions differing in the mean only), these two would be equivalent.

The transformations described in the last two methods result in mappings of the mean (Ho et al. 2012):

$$\text{bias correction, } \mu_{\text{of}} = \mu_{\text{oh}} + \frac{\sigma_{\text{oh}}}{\sigma_{\text{mh}}}(\mu_{\text{mf}} - \mu_{\text{mh}});$$

$$\text{change, } \mu_{\text{of}} = \mu_{\text{mf}} + \frac{\sigma_{\text{mf}}}{\sigma_{\text{mh}}}(\mu_{\text{oh}} - \mu_{\text{mh}}).$$

In both cases, the standard deviation maps as $\sigma_{\text{of}} = (\sigma_{\text{oh}}\sigma_{\text{mf}})/\sigma_{\text{mh}}$. For constant and unbiased variance ($\sigma_{\text{mf}} = \sigma_{\text{mh}} = \sigma_{\text{oh}}$) the two cases are the same, consistent with the linearity argument just made. While these results are general, the mean and variance only fully describe the distribution and thus any changes in it for Gaussian distributions.

We note at this point that calibration implies an assumption that the model–observation discrepancy is due to model error rather than observational error. This may be a poor assumption in regions of high observational uncertainty. Furthermore, bias correction can only lead to useful projections if the model is able to plausibly simulate climate change (Maraun 2016) and therefore the processes that govern that change. Very large biases may be an indication of fundamental shortcomings in physical processes in the model, indicating that it does not produce these required plausible projections of change.

c. Analytical model

If temperatures are assumed to be a sample from a known statistical distribution, it may be possible to derive an analytical distribution for the expected sum of HDD or CDD over a defined period. Let daily temperature X have probability distribution function $f_X(x)$. Denoting the daily value of HDD by the random variable Y , $Y = g(X) = (T_H - X)H(T_H - X)$, the expected value of HDD is given by

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} (T_H - x)H(T_H - x)f_X(x)dx. \end{aligned} \quad (3)$$

Further assuming that X is normally distributed $X \sim N(\mu, \sigma^2)$ (the validity of this assumption will be discussed later), it is possible to evaluate this integral (derivation in the appendix) arriving at the result:

$$\mathbb{E}(Y) = \frac{(T_H - \mu)}{2} \left[1 + \operatorname{erf}\left(\frac{T_H - \mu}{\sigma\sqrt{2}}\right) \right] + \frac{\sigma}{\sqrt{2\pi}} e^{-[(T_H - \mu)^2/2\sigma^2]} \quad (4)$$

for HDD and

$$\mathbb{E}(Y) = \frac{(\mu - T_C)}{2} \left[1 + \operatorname{erf}\left(\frac{\mu - T_C}{\sigma\sqrt{2}}\right) \right] + \frac{\sigma}{\sqrt{2\pi}} e^{-[(\mu - T_C)^2/2\sigma^2]} \quad (5)$$

for CDD, where erf is the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy. \quad (6)$$

To obtain the expected sum of HDD or CDD in a period of length n (e.g., the expected value of total HDD in a season), these expected values are multiplied by n . This derivation assumes X to be stationary but does not assume independence.

It would also be possible to derive an expression for the variance of daily degree-days. However, for useful applications, the variable of interest would be the interannual variance of the seasonal sum of degree-days. For example, a cool climate with negligible average CDD might have individual hot years with nonnegligible CDD, resulting in a large, interannual variance that would be important to be aware of. In contrast to the expression for the expectation (mean), where one can simply scale by n , an expression for this interannual variance would have to allow for the nonindependence of X ; it would only be valid if it appropriately accounted for the dependence structure of daily temperatures. In the current study, we are interested in climatological degree-days rather than interannual variability and therefore do not pursue this further.

PROPERTIES AND IMPLICATIONS OF ANALYTICAL EXPRESSION

Equations (4) and (5) are nonlinear in μ and are not a function of μ alone. This is as expected because the index $g(X)$ is nonlinear (Weisheimer and Palmer 2014). This nonlinearity will have implications for treatment of biases. We therefore explore the behavior of Eqs. (4) and (5) by looking at their derivatives. The first derivative is

$$\begin{aligned} \frac{\partial}{\partial \mu} [\mathbb{E}(Y)] &= -\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{T_H - \mu}{\sigma\sqrt{2}}\right) \right] \text{ for HDD or any below-threshold DD;} \\ \frac{\partial}{\partial \mu} [\mathbb{E}(Y)] &= +\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{\mu - T_C}{\sigma\sqrt{2}}\right) \right] \text{ for CDD or any above-threshold DD.} \end{aligned} \quad (7)$$

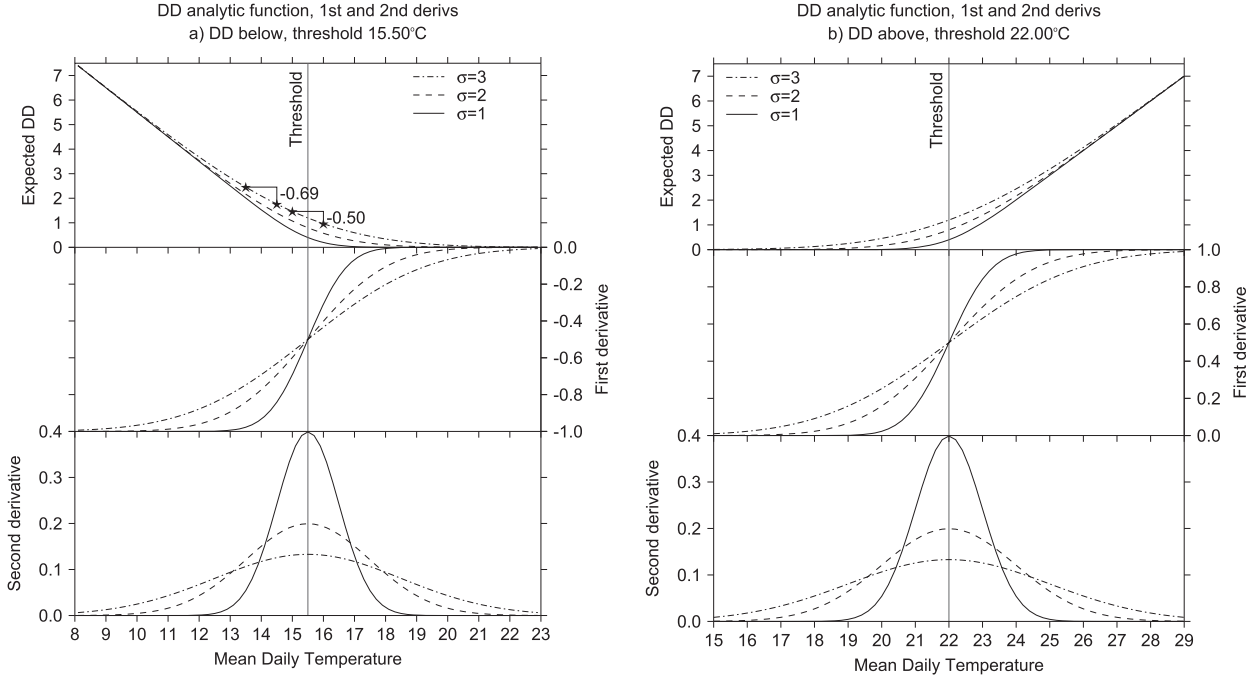


FIG. 1. The behavior of the expected value of daily (a) HDD and (b) CDD, following Eqs. (4) and (5). The panels show (top) expected value, (middle) its first derivative, and (bottom) its second derivative, with respect to the mean of the underlying temperature distribution. The vertical gray line shows the threshold. Different line types show the effect of changes in the standard deviation, as shown in the legend. The step changes in (a) demonstrate the implications of the function nonlinearity (see text).

By definition of the error function, the second derivative is (for a general threshold T_A)

$$\frac{\partial^2}{\partial \mu^2} [\mathbb{E}(Y)] = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(T_A - \mu)^2}{2\sigma^2}} \equiv \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu - T_A)^2}{2\sigma^2}}. \quad (8)$$

These functions are shown in Fig. 1. We will focus initially on the “below threshold” case HDD. For cold climates, a long way below the threshold, expected HDD are approximately a linear function of μ . This is evident from the graph of the function itself (Fig. 1a, top) and the fact that the second derivative is approximately equal to 0 (Fig. 1a, bottom). The function is approximately independent of σ far from the threshold, where the functions for different values of sigma converge (Fig. 1). However, although linear, the sensitivity of expected HDD to μ and so to any mean bias is greatest far below the threshold, where the absolute value of the first derivative is maximum. In contrast, for warm climates (μ far above the threshold), expected HDD tend to zero. Equivalent behavior occurs for CDD or any degree-day above a threshold (Fig. 1b).

When the mean is near to the threshold temperature, the function is nonlinear (Fig. 1). The nonlinearity is greatest when the second derivative is largest; this

occurs when the mean is equal to the threshold temperature (bottom panel).

In Fig. 1a (top), we illustrate the implications in the case when only μ is biased and only μ changes (i.e., σ is unbiased and does not change). The left-hand “step” represents observations, in which a change in the mean of 1°C results in a change of -0.69 HDD. (This is a daily value, equivalent to approximately -20 HDD per month or -250 HDD per year.) The central step represents a model, in which the same change in the mean temperature results in a change of only -0.5 HDD. Therefore, assuming that it is valid to assume that the underlying change in the mean temperature is represented correctly in the model, applying the change in HDD from the model to the observations is not appropriate. This arises because nonlinearity means that $\mathbb{E}[g(X_{or})] \neq \{\mathbb{E}[g(X_{oh})] + \mathbb{E}[g(X_{mr})] - \mathbb{E}[g(X_{mh})]\}$, but the additive index correction method assumes that the two sides of this expression are in fact equal.

d. Gridded data

1) REGIONAL CLIMATE MODEL DATA

To evaluate future change in degree-days, we use daily data at high spatial resolution (25-km grid) over the United Kingdom from the Hadley Centre regional

model 11-member perturbed physics ensemble (PPE), HadRM3-PPE (Hadley Centre for Climate Prediction and Research 2008; Murphy et al. 2010). A PPE is produced by adjusting certain parameters within the model that control physical processes smaller than the grid scale and that are not precisely known. This differs from an initial condition ensemble, in which the only difference between runs is a small perturbation to initial conditions. HadRM3-PPE was produced as part of the U.K. Climate Projections 2009 (UKCP09) to explore the uncertainty in climate projections because of uncertainty in physical processes (sections 3.1, 3.2, and 5 of Murphy et al. 2010). Many parameters are perturbed together such that it is not possible to attribute differences to specific parameters. The members have labels *afgcx* (the unperturbed member) and *afix_* for the 10 perturbed members, with varying final character.

HadRM3-PPE was produced by dynamically downscaling HadCM3-PPE (Murphy et al. 2010), an ensemble simulation using the HadCM3 coupled model. HadCM3-PPE was run from 1860 to 2100, forced with historical emissions for 1860–2000 and with emissions scenario A1B (Nakicenovic and Swart 2000) from 2000 onward (Murphy et al. 2010). The downscaling was performed for the period 1950–2099. Collins et al. (2011) and Lambert et al. (2013) contain further discussion of the parameter perturbations.¹ The configuration and performance of HadCM3 was discussed in Gordon et al. (2000).

In contrast to UKCP09 standard output, we use the more recent 1981–2010 as the climate baseline period.² Positive trends in average temperature have been observed since 1960 in all seasons and are statistically significant at the 95% level in all seasons except winter (e.g., Sniffer 2014). Therefore, the 1981–2010 climatology is generally warmer than that of 1961–90. For future change, changes are evaluated for the period 2040–69 relative to the baseline.

We are not aware of existing evaluation of temperature biases in HadRM3-PPE in the literature. Brown et al. (2010) mention that snow biases in western Scotland in the ensemble are likely to be related to known cold biases here. Other evaluation of HadRM3-PPE

biases focuses on precipitation (e.g., Sanderson et al. 2012).

2) OBSERVATIONAL DATA

We evaluate biases in the climate model against the European gridded observational dataset E-OBS, which was generated as part of the ENSEMBLES project (Haylock et al. 2008). The data are available at daily resolution on the same 25-km rotated grid as HadRM3-PPE. E-OBS is updated approximately every 6 months with updated station series and, where available, updates to the station network. This project uses version 12.0, which was released in October 2015.

Over the United Kingdom, there are known biases in E-OBS relative to a 5-km daily dataset produced by the Met Office (Hofstra et al. 2009; Perry and Hollis 2005), hereafter referred to as UKMO. The evaluation in Hofstra et al. (2009) was performed for an earlier version of E-OBS, but updates to the station network used over the United Kingdom have been limited, and so it is assumed that their conclusions are largely unchanged. Hofstra et al. (2009) reported root-mean-square errors in daily temperatures of up to 1.4°C in northwest Scotland (their Fig. 4), the region of most disagreement between datasets, and 0.7°C when averaged over the United Kingdom.

The UKMO climatology is available for 1961–90 on a 25-km grid. Over this period, for a U.K. average, we find that E-OBS is warmer than UKMO (not shown); in the annual mean, MAM, SON, and DJF, the difference is approximately 0.1°C, while in JJA it is approximately 0.2°C. It will be shown later that HadRM3-PPE biases relative to E-OBS are larger than this in general. We therefore use E-OBS as UKMO is not available at daily resolution on the 25-km grid. The discrepancies should be born in mind. In particular, E-OBS is over 1°C warmer than UKMO in northern Scotland in DJF and warmer in general over Scotland.

3) SPATIAL AVERAGING

We present results for two U.K. cities, London and Glasgow. Temperature in each city is calculated as the mean over every grid point within a circle of radius 0.44°, equivalent to two grid points, about its central coordinates. Too broad an averaging region could smooth out the extremes of temperature that might lead to HDD or CDD, but it is not advisable to use single-gridpoint information from climate models because of the noise. The urban heat island effect—whereby an urban area is hotter than its rural surrounds—is not represented in the regional model and only to a limited extent in E-OBS. Projections of future absolute values should therefore be interpreted with care because the

¹ Further information on the ensemble, including the climate sensitivities of the corresponding global slab ocean models, can be found online (at https://badc.nerc.ac.uk/artefacts/badc_datadocs/hadrm3-ppe-uk/index_140211.html).

² The period 1981–2010 is used for calculation of anomalies in the annual U.K. Met Office “State of the Climate” report (Kendon et al. 2016). However, the official World Meteorological Organization climatological normal reference period will remain as 1961–90 until data for the period 1991–2020 are available (Trewin 2007).

TABLE 1. U.K.-mean annual biases in individual model members. Biases are calculated at each grid point and then spatially averaged. Shown for period 1981–2010, the focus of the study, and for 1961–90 against two observational datasets as a measure of robustness. Headers -c, -h, etc., refer to member afxc, afhx, and so on. Subseasonal variance is the variance of anomalies from the seasonal cycle (see text, section 3a).

	1 (afgcx)	2 (afixa)	3 (-c)	4 (-h)	5 (-i)	6 (-j)	7 (-k)	8 (-l)	9 (-m)	10 (-o)	11 (-q)
1981–2010											
Mean bias (member-E-OBS)	−0.76	−0.89	−1.34	−0.90	−0.33	−0.32	−0.33	−0.56	0.02	−0.41	0.07
Subseasonal variance ratio $\frac{\text{member}}{\text{E} - \text{OBS}}$	1.09	1.46	1.47	1.01	1.14	1.30	0.88	1.24	1.07	0.90	1.28
1961–90											
Mean bias (member-UKMO)	−0.61	−0.81	−1.14	−0.77	−0.02	−0.15	−0.02	−0.48	0.06	−0.35	0.23
Mean bias (member-E-OBS)	−0.72	−0.91	−1.26	−0.89	−0.14	−0.26	−0.14	−0.59	−0.07	−0.47	0.10
Subseasonal variance ratio $\frac{\text{member}}{\text{E} - \text{OBS}}$	1.53	2.04	1.33	1.27	1.08	1.43	0.91	1.13	1.60	1.01	1.35

bias correction will not fully account for it (since it is not fully represented in the observations). Moreover, any uncertainty over change in the strength of the urban heat island would add to the uncertainty in projections of degree-day indices.

3. Results

a. Biases in the daily temperature distribution

So that we can apply the above analytical result to a realistic case, we quantify the biases in the HadRM3-PPE ensemble members relative to E-OBS for the United Kingdom. As discussed previously, the different members in a PPE are different model configurations and so differences between them represent true differences in the simulated climate as well as differences in sampling. In contrast, in an initial condition ensemble the differences are only of sampling.

Accurate representation of temperature indices depends on simulation of the mean temperature and its seasonal cycle as well as subseasonal variance (weather noise). Therefore, we present an analysis of mean temperature in standard meteorological seasons winter (DJF) and summer (JJA), the seasonal cycle at individual locations, and subseasonal variance.

To compare data from the normal calendar in E-OBS with the 360-day calendar in HadRM3-PPE, leap days are discarded from the E-OBS dataset. Calendar months in E-OBS are then compared with 30-day months in HadRM3-PPE. To compare seasonal cycles, the cycle is calculated on its native calendar, and then the E-OBS cycle is scaled for plotting.

The annual-mean, regional-mean bias in each ensemble member for the period 1981–2010 (Table 1) ranges from -1.34°C (member afxc) to $+0.07^{\circ}\text{C}$ (member afxq); 9 of 11 members have a cold bias. As a simple measure of the robustness of these biases, the

biases calculated in the period 1961–90 are also shown; biases in this earlier period are similar but with a tendency to be slightly warmer (equivalently weaker cold bias) than in the later period. In this earlier period, performing the same comparison against UKMO (Table 1) reveals biases that are weaker by approximately 0.1°C . In conclusion then, the ensemble simulations are biased cold in general, but the cold bias against “truth” could be slightly exaggerated when evaluated against E-OBS.

The annual-mean, regional-mean bias in mean temperature masks regional and seasonal behavior, which varies between ensemble members. There are shared characteristics; cold biases tend to be strongest in western Scotland, where cold biases are found in all members and all seasons. Figure 2 shows three representative members as an example; the unperturbed member afgcx as well as afxc and afxj. Member afxc has the strongest cold bias of all members and is biased cold everywhere and in both seasons. Member afxj has the strongest warm biases in the south in summer but has cold biases in winter.

To examine the seasonal cycle, we calculate the leading Fourier components (capturing the mean and the annual and half-annual frequencies) in each dataset. Figure 3 shows the seasonally varying observed and simulated temperature for London. London is discussed because, first, it demonstrates a variety of behaviors between the different members, and, second, as a large population center, it is important for total energy demand. In summer, members afxi, afxj, afxk, and afxq and to a lesser extent afxl and afxm have a warm bias (Fig. 3; July average warm bias in these members ranges from 0.45° to 1.44°C). In winter, members afxi, afxj, afxk, and afxq have a small (less than 1°) warm bias, but these members have no or cold bias in spring and autumn. Members afxc, afxj, afxk, and afxq have previously been shown (Sanderson et al. 2012, their supplementary information)

1981-2010 CLIMATOLOGY, AND SELECTED ENSEMBLE MEMBER BIAS

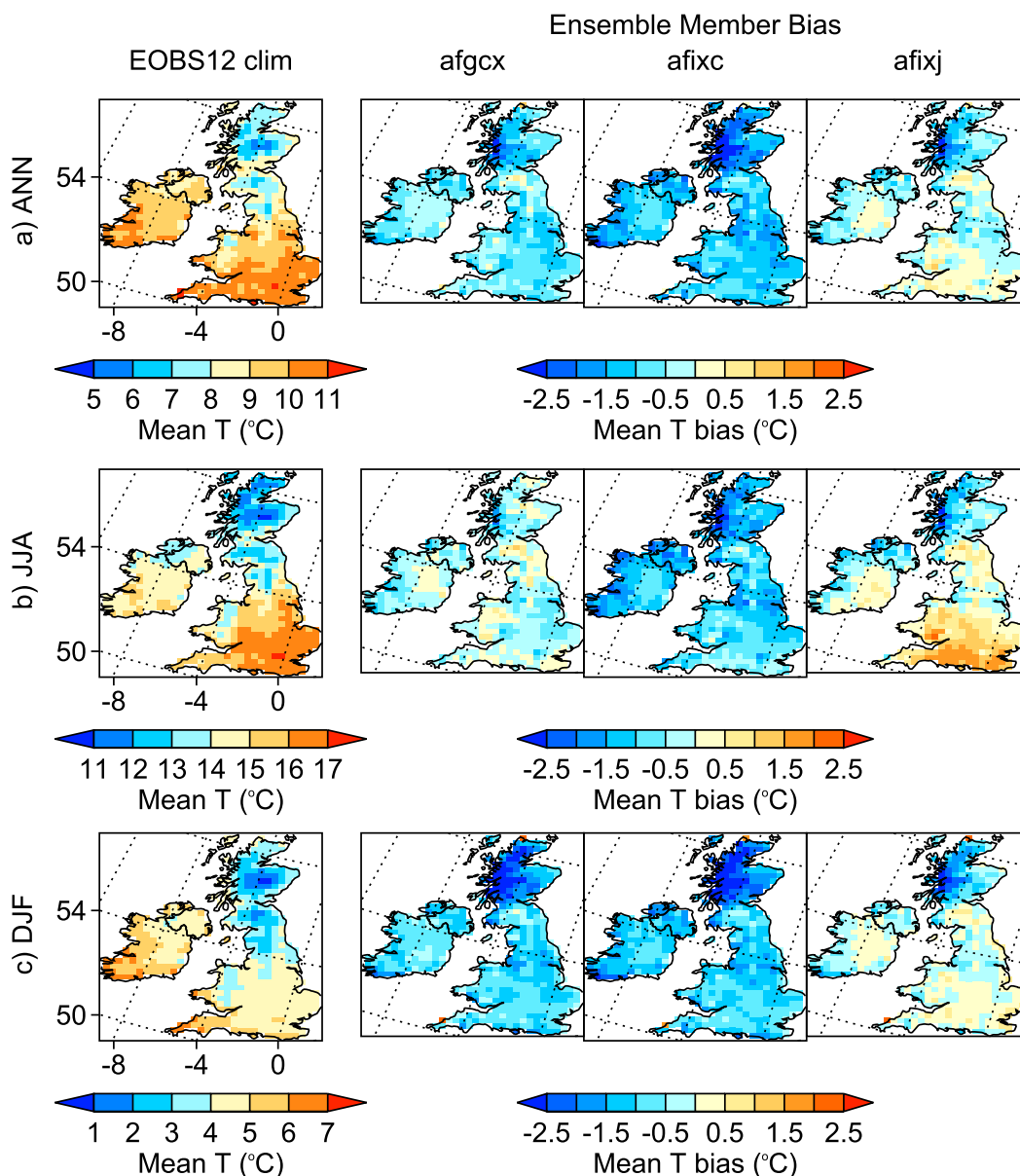


FIG. 2. E-OBS12 temperature climatology for the period 1981–2010, and biases in three selected members of the ensemble HadRM3-PPE. As labeled, results are for (a) annual (ANN), (b) summer (JJA), and (c) winter (DJF); DJF climatology is for December 1980–February 2010.

to be too dry in summer. Three of these four members are those just discussed as having a warm bias. This suggests that feedbacks caused by drying and lack of soil moisture may help explain the too-hot summer bias. The remaining member, afixc, is biased cold in all months. Therefore, these feedback processes may be insufficient to produce enough warming to overcome the general cold bias.

To quantify subseasonal variance (see start of this section), we calculate the variance of anomalies from the

calculated seasonal cycle. The regional-mean proportional bias (model/observed) of this quantity is shown in the second row of Table 1. The spatial structure varies (not shown); however, there is a general tendency for the models to have greater subseasonal variance than the observations.

Temperature distributions for individual months, evaluated in London (Fig. 4), suggest that the Gaussian distribution is a good assumption for within-month

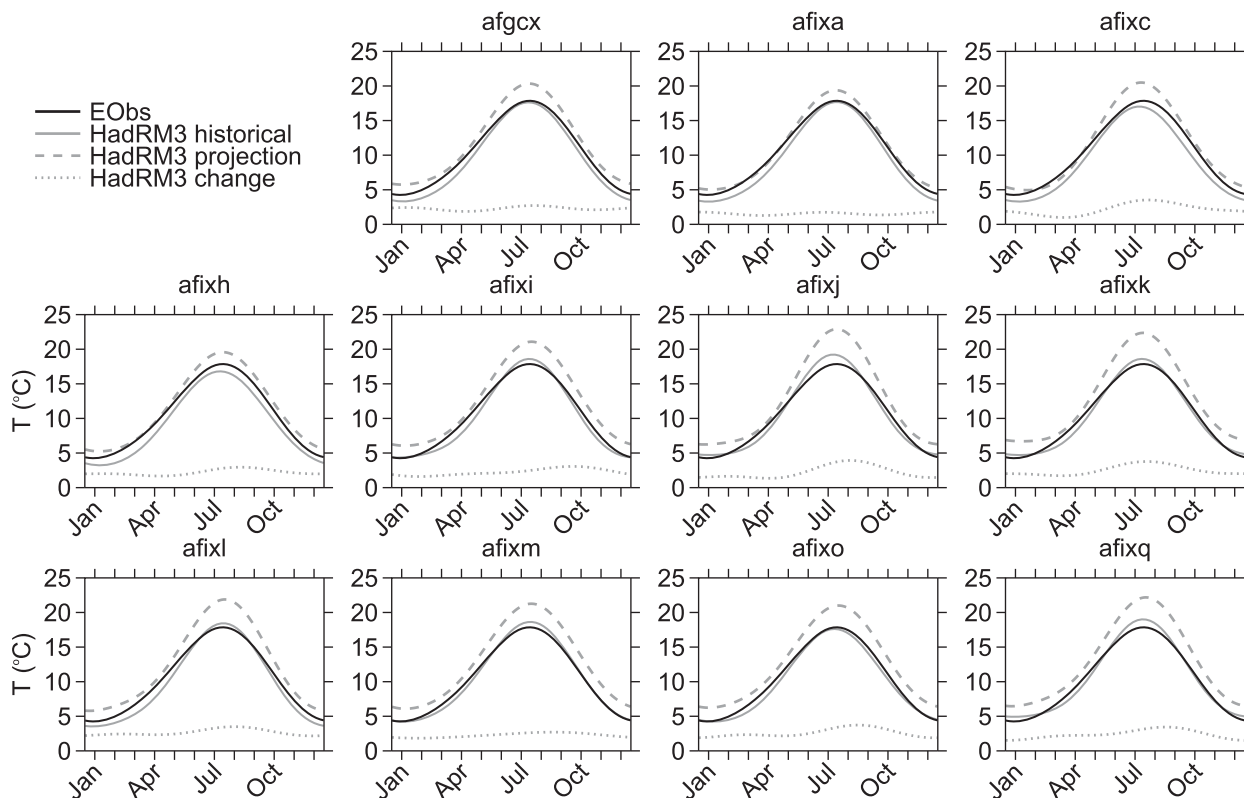


FIG. 3. Seasonal cycle (leading Fourier components; see text) of temperatures in London. E-OBS (1981–2010, black solid lines), modeled past (1981–2010, gray dashed lines), modeled future (2040–69, gray dashed lines), and difference between modeled past and modeled future (gray dotted lines). Each panel shows a different ensemble member.

temperatures in the observations and in the spring and autumn in the model. There is evidence of positive skew in simulated data in July (found in most ensemble members; not shown) and of deviations from the Gaussian in January (found in about half the ensemble members; not shown).

Since biases are seasonally varying, analysis in the rest of the study is performed on a month-by-month case.

Therefore, we do not make further use of the Fourier component–based seasonal cycles; instead, we use monthly means and the variance of anomalies from those monthly means.

b. Test cases

We now explore the implications of the different calibration strategies (section 2b) for DD projections

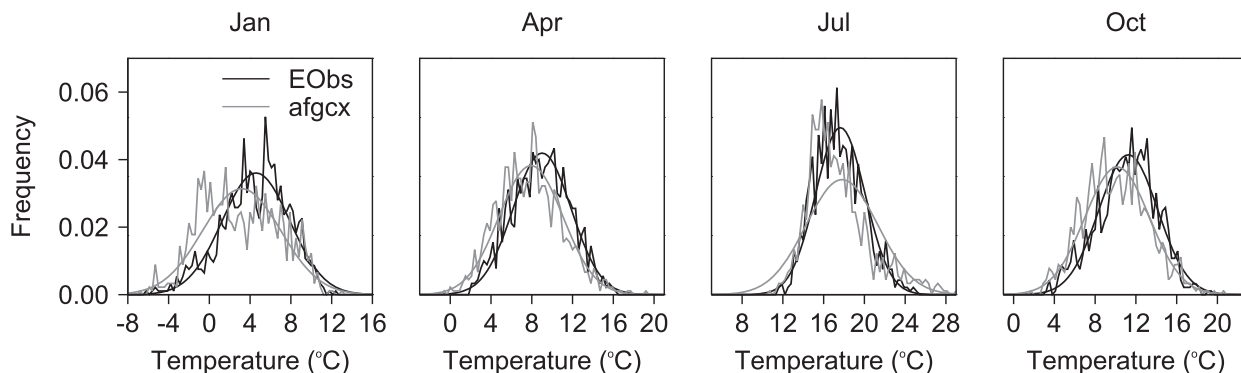


FIG. 4. The sampled distribution of temperatures in London, and a Gaussian approximation to the distribution, in E-OBS (black) and the unperturbed member afgcx (gray). This is shown for four individual months (January, April, July, and October) to sample the four seasons.

TABLE 2. Test case HDD and CDD projections using the analytical method of sections 2c [Eqs. (4) and (5)] and calibration strategies of section 2b for three representative cases. Index I from the main text is replaced by H and C . Time periods are 1981–2010 baseline, 2040–69 future. Biases and change based on unperturbed member afgx, other than in London July case where member afij is used to demonstrate impact of large bias. For subscript notation ($_{oh}$, etc.) and details of method, see sections 2b and 3b.

Parameter	Description	1) Glasgow, January	2) Glasgow, July	3) London, July
μ_{oh}	Observed mean T	3.2	14.6	17.6
σ_{oh}	Observed standard deviation T	3.0	2.1	2.4
$\varepsilon\mu$	Model bias, mean T	−2.2	−0.3	1.4
$\varepsilon\sigma$	Proportional model bias, standard deviation T	1.2	1.1	1.8
$\Delta\mu$	Model change, mean T	2.7	1.6	3.3
$\Delta\sigma$	Proportional model change, standard deviation T	0.9	0.9	1.0
H_{oh}	HDD calculated from μ_{oh} , σ_{oh}	368	41.1	7.5
H_{mh}	HDD calculated from μ_{mh} , σ_{mh}	435	50.6	15.4
HDD projections:				
H_{mf}	Calculated from μ_{mf} , σ_{mf}	354	21.1	3.8
$H_{\Delta I}$	Additive index correction ($=H_{oh} + H_{mf} - H_{mh}$)	288	11.6	−4.2
$H_{\times I}$	Proportional index correction ($=H_{oh}H_{mf}/H_{mh}$)	300	17.1	1.8
$H_{\Delta T}$	Underlying change method	291	14.7	0.4
H_{eT}	Underlying bias correction method	299	16.3	1.8
C_{oh}	CDD calculated from μ_{oh} , σ_{oh}			1.0
C_{mh}	CDD calculated from μ_{mh} , σ_{mh}			19.4
CDD projections:				
C_{mf}	Calculated from μ_{mf} , σ_{mf}			60.1
$C_{\Delta I}$	Additive index correction	—	—	41.7
$C_{\times I}$	Proportional index correction	—	—	3.2
$C_{\Delta T}$	Underlying change method	—	—	16.7
C_{eT}	Underlying bias correction method	—	—	6.3

in three cases chosen to represent different climates: Glasgow in January, Glasgow in July, and London in July. Within-month daily temperatures are assumed to follow a Gaussian distribution, the validity of which is discussed above. Equations (4) and (5) with $n = 30$ days = 1 month [n is a scaling factor for monthly or seasonal totals; see text after Eq. (5)] are used to calculate the expected monthly value of DD. The parameter inputs to these equations are the estimated statistical properties of the different temperature distributions; for index correction methods, this is the sample mean μ and variance σ^2 from observations and from historical and future simulations, while for the underlying calibration methods these are calculated according to the transformations in section 2b. The following discussion is centered on the results in Table 2. The general index I is replaced by H (heating degree-days) and C (cooling degree-days). Therefore, the results from different methods for heating degree-days are denoted $H_{\Delta I}$, $H_{\times I}$, and so on. (For subscript definitions and descriptions, see section 2b.)

1) CASE 1: HDD WHEN $\mu \ll T_H$

As shown in section 2c, HDD is approximately linear in μ and independent of σ for sufficiently small μ . This follows intuitively; when the mean temperature is

sufficiently cold, every day is colder than the threshold temperature. Therefore, the indicator function that introduces nonlinearity is irrelevant and the problem reduces to a linear one. Case 1 of Table 2 (Glasgow in January) was chosen as a “cold climate” case to illustrate this point. However, although the historical mean temperature is far below the threshold, the large projected change brings temperature into the regime where nonlinearity starts to be important, so the different strategies produce slightly different projections. Even so, these differences are small relative to the total number of HDD. For slightly colder climates, or smaller projected changes, the projections from different methods are indistinguishable (not shown).

2) CASE 2: HDD WHEN $\mu \approx T_H$

In a second case, μ lies just below the 15.5°C threshold. This applies for much of the United Kingdom in the summer months. Here, we consider Glasgow in July (case 2; Table 2). The projected value of HDD in this case differs more according to method, as expected from the nonlinearity of the function in this domain. The index correction method yields a projection of 11.6 HDD, the change method yields a value of 14.7 HDD, and the bias correction method yields a value of

16.3 HDD. This represents a large proportional difference between methods, even though, in this climate, differences in July HDD represent a small difference relative to the annual total that is dominated by winter HDD. The effect of these methods on annual total HDD (and CDD) projections will be investigated in [section 3c](#).

3) CASE 3: HDD WHEN $\mu > T_H$

Finally, we consider London in July in member afixj. Here, there is a positive bias in both μ and σ (case 3; [Table 2](#)). Despite the warm bias, the positive variance bias means there are too many HDD. Large future warming results in a large reduction in HDD. Applying the additive change in the index $[7.5 + (3.8-15.4)]$ gives a projected value of -4.2 HDD. This negative value violates the definition of HDD and so must be incorrect, thereby demonstrating that the additive index correction cannot give a valid answer. The proportional change method gives a result of 1.8 HDD, while the underlying correction methods give values of 1.8 HDD (bias correction) or 0.4 HDD (change method).

4) CASE 4: CDD

We now consider CDD. Given the threshold $T_C = 22.0^\circ\text{C}$, all months in all U.K. grid boxes have $\mu < T_C$ in current climate and often $\mu \ll T_C$. Therefore, days above the threshold are very rare, or equivalently CDD only arise from anomalously warm days. In this case, Eq. (5) is very nonlinear in μ ([Fig. 1](#)). The bottom section of [Table 2](#) explores projected CDD for London in July, again for member afixj (see case 3). Because the model is biased hot, the modeled values of μ and σ give too many CDD, that is, $C_{mh} > C_{oh}$. Future warming means that CDD will increase. The additive index correction method gives a projection of 41.7 CDD. Other methodologies give much smaller, although differing, projected values: 3.2, 16.7, and 6.3 CDD. This large difference in results is due to the nonlinearity of the index and the large bias in both mean and variance. From our earlier discussion, we may conclude that the additive index method is inappropriate in this nonlinear case, but it is unclear which of the other three methods is most appropriate.

Different ensemble members also give different answers after calibration; this behavior is discussed further in [section 3c](#).

As mentioned in [section 2c](#), our model accounts only for the climatological mean and not for interannual variability. In a climate such as that of the present-day United Kingdom, CDD are largely driven by extreme events in individual years, such that the climatological mean is not necessarily the most useful metric.

5) IMPLICATIONS

Case 1 (HDD in a cold climate) demonstrates that in some cases, relative differences between a simple index method and a more complex methodology are small. Therefore, a user interested in projecting such an index could expect calculating the change in the index from model output and adding it to the observed index to provide a reasonable estimate of future HDD, saving time and cost. (This assumes of course that the user trusted the climate model's projected changes in the underlying temperatures.) Furthermore, returning to the equation for the index demonstrates that the additive index correction and not the proportional index correction is correct in the linear case.

In other cases, the examples above demonstrate that calculating changes in the index alone will not provide a reliable projection. It is not obvious which of the two underlying methods is appropriate. Users should explore the choice of methodology as a source of uncertainty, as previously proposed in agricultural studies ([Ruiz-Ramos et al. 2016](#)) and in [Ho et al. \(2012\)](#). The next section explores these uncertainties, and their magnitude, for U.K. degree-day projections.

In the above we have discussed the difference between methods as a difference relative to the absolute number of DD. Our motivation for doing so is that DD enter energy models as a multiplicative factor ([Isaac and Van Vuuren 2009](#)) so that it is relative differences that are important. The absolute differences between methods are largest in case 1 (although as commented, they do reduce to zero for slightly colder climates). This is discussed further in the conclusions.

One might ask at what point nonlinearity becomes important. The second derivative of the expectation function is a Gaussian [Eq. (8)], its scale determined by σ . Therefore, one can determine how many standard deviations from the mean bring the nonlinearity function below any predefined threshold. The challenge, then, is to define such a threshold. This may well be user specific, but, in general, [Fig. 1](#) suggests that nonlinearity is certainly unimportant when the mean temperature is (and remains under climate change) more than 3-sigma away from the threshold.

Finally, the conclusions above rely on DD values derived assuming that the distributions are Gaussian. Given that there is some evidence that the Gaussian assumption breaks down in summer and winter for some ensemble members ([section 3a](#)), the expected DD derived using these expressions will be a biased estimate, particularly in the CDD case. Therefore, the values are illustrative only, but the chain of reasoning that seeks to establish the state of climate for which results are

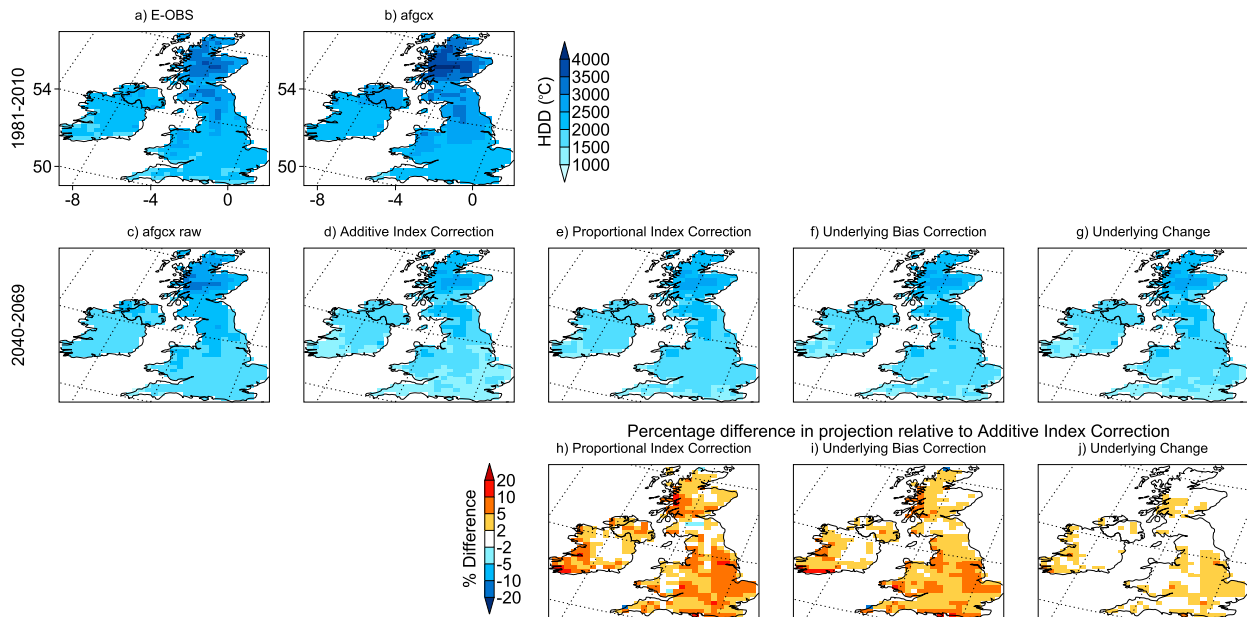


FIG. 5. The effect of different bias correction strategies in the unperturbed member afgcx for the period 2040–69. Historical HDD in (a) observations and (b) model. (c)–(g) Projected future HDD according to different methodologies. (h)–(j) The percentage difference between projected values of HDD for different strategies, relative to the additive index method.

particularly sensitive to the bias correction method remains valid and could be extended with care to distributions with, for example, pronounced skew.

c. Projections of HDD and CDD

We now apply the four calibration methods discussed earlier to climate model output to calculate projections of future annual HDD and CDD. In the case of calibration strategies for the underlying data, transformations are now applied to the full distribution, but correcting for mean and variance only. A data file containing spatial projections on a month-by-month basis is available (Holmes 2017).

It is generally accepted that some calibration or bias correction is necessary. Therefore, for each method we present the projected value and the percentage difference between this value and the simple “additive index bias correction” projection. Proportional change and difference is used for the reason discussed above: that DD are generally used as multiplicative factors in energy models (Isaac and Van Vuuren 2009).

Figures 5 and 6 show maps of projections using different methods for a single ensemble member. For HDD, the methodologies that treat the underlying distribution tend to project higher future values than the index correction methods, with differences of over 5% at many grid points in the unperturbed member (Figs. 5h,i,j). This is representative of most model

members. Consistent with the case studies above, then, the relatively cold U.K. climate (and therefore approximate linearity of HDD in mean temperature) means that the effect of calibration choice is relatively small. However, users sensitive to small margins (such as 5%) or changes in particular seasons may still need to be concerned about these differences.

On the other hand, for the CDD method choice has a large relative impact on the magnitude of projections. This is evident in Fig. 6. While the absolute values of CDD involved are small, such large differences in projections could be crucial to stakeholders deciding whether to invest in cooling technology.

Figure 7 shows the projections averaged over the United Kingdom, Glasgow, and London for each ensemble member. As well as the calibrated projections (colored), raw model output is also shown (dashed black: historical; solid black: future). The ensemble members are sorted by their historical bias in each metric. For all members except 9:afixm and 11:afixq, all calibration methods produce a large reduction in projected HDD relative to the raw model output. This is particularly true for members 1 to 4 (right of plot, Fig. 7a). This is consistent with their large historical cold biases (Table 1). In general, the methods applied to underlying data project higher future HDD than the simple additive index correction method (blue) does (Fig. 7a). Similar results are found for Glasgow (Fig. 7b), although the effect, and therefore importance,

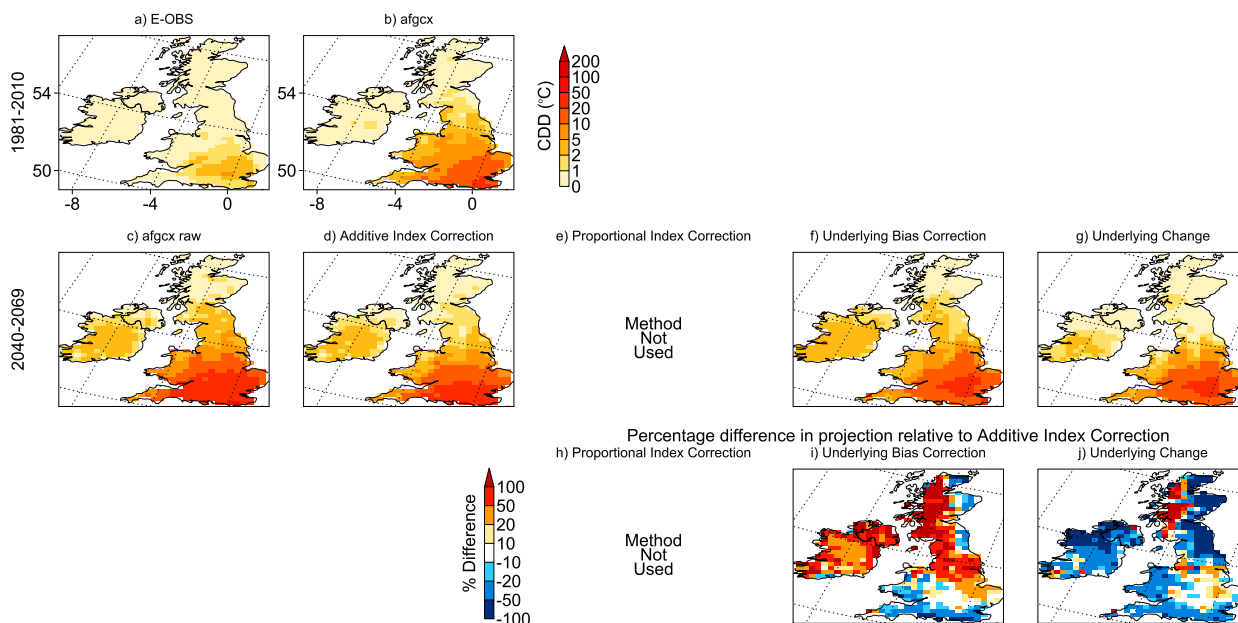


FIG. 6. As in Fig. 5, but for CDD. Proportional index correction method not appropriate because of the historical value of 0 in some months.

of calibration is even larger. This is consistent with larger biases relative to the national case.

It has been suggested that bias correction/calibration reduces uncertainty in impacts' projections (e.g., Ruiz-Ramos et al. 2016). Figure 7a shows that for HDD, this is indeed the case; the spread across all methods across all members (colored lines) is less than the spread in the raw output (black line). However, the calibration methods introduce a new source of uncertainty, especially for members with large historical bias (right of Fig. 7a). For this ensemble, the introduced “calibration uncertainty” is approximately equivalent to the uncertainty from the ensemble spread; the range in values in any one method (colored line) is approximately equal to the spread between colored lines for any one member. This is quantified in Table 3.

For all members, additive index correction reduces the projected value of CDD relative to the raw model output (Figs. 7c,d, blue vs black line). Projections from both the underlying change method and the underlying bias correction method (red and yellow) are lower than that from the simple additive index correction method (blue) in most members. The differences in the projections from underlying methodologies, which are large in some members (e.g., afgcx, London; Fig. 7d), must be attributable to changes or biases in the variance (Ho et al. 2012).

Again, the uncertainty in CDD associated with the choice of calibration methods is equivalent in magnitude to that associated with the ensemble spread (Table 3). This

is true even if the additive index correction is, consistent with our earlier discussion, discounted and only the two methods applied to underlying data are considered.

The uncertainty in the HadRM3-PPE ensemble used here is not a full measure of the uncertainty in climate change projections because of the model physics or initial condition uncertainty. The calibration uncertainty from these methods (which is not the full calibration uncertainty, as these are only a selection of relatively simple methods) would therefore reduce in relative importance when compared to a broader assessment of uncertainty in climate change, such as that used in UKCP09.

4. Conclusions

In this paper, we have analyzed degree-day (DD) indices using an analytical tool and climate model output. We have developed an analytical tool that can be used to explore the effect of different calibration strategies without detailed analysis of daily climate model output. This work has extended concepts previously largely applied in agricultural studies regarding the unreliable results that can arise from applying certain bias correction strategies. While some previous studies have focused on the importance of bias correcting variance as well as mean, we demonstrate that problems with simple bias correction methods may arise even for mean biases only. Our main conclusion is that sufficiently far from the threshold used to define DD, the DD index is

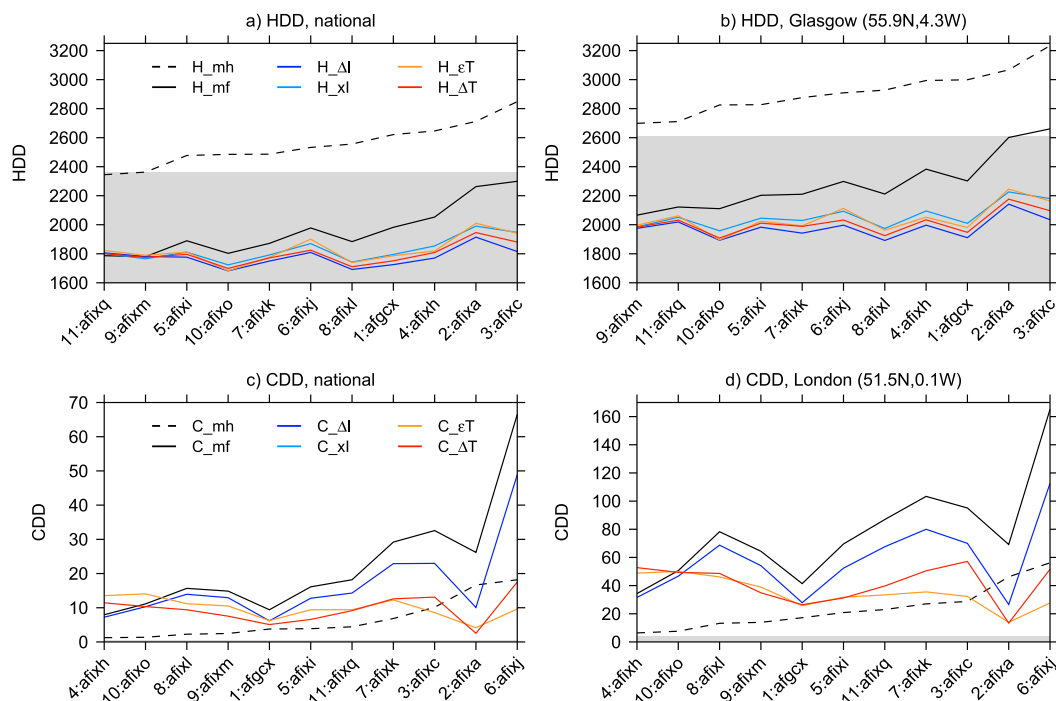


FIG. 7. Projected DD (solid lines) for 2040–69 from the 11 ensemble members: (a) HDD, U.K. national average; (b) HDD, Glasgow; (c) CDD, U.K. national average; and (d) CDD, London. The top of the gray background indicates the observational historic (1981–2010) value, and the black dashed line shows the modeled historic value. Models are sorted by their historical bias. Black solid lines show the uncalibrated model projection, while each colored solid line shows the result from a calibrated projection. The terms H_{mh} and H_{mf} are raw model historical and future output, and in other cases subscripts as in section 2b are typeset, for example, $H_{\Delta I} = H_{\Delta I}$.

approximately linear in the mean temperature, and simple index correction may be appropriate. Close to the threshold, large uncertainty is introduced by different calibration methodologies. As a rough approximation, for Gaussian distributions, the linear regime applies when the mean temperature is more than three standard deviations away from the threshold. Alternatively, more simply, the linear regime can be assumed to apply for HDD if almost all observed daily temperatures are below the threshold—equivalently, when the maximum observed daily mean temperature is below or only just above the threshold.

Consistent with the above, projections for the United Kingdom suggest rather different conclusions for HDD and CDD. For HDD, even a simple index bias correction method largely removes the spread in member projections introduced by model bias. However, for members with nonnegligible bias, the choice of method introduces new uncertainty. For CDD on the other hand, the same simple index correction method is certainly inadequate, giving much larger projected values than more advanced methods applied to the underlying temperature distribution. This

suggests that studies such as Zubler et al. (2014) and Erhardt (2015) should indeed be extended to take account of the effects of nonlinearity. The maximum plausible CDD projection (discounting the simple index correction method) we have found for the U.K. average in 2040–69 is under 20 CDD, and the minimum plausible (any correction method) HDD projection is 1710 HDD (Figs. 7a,c). This implies that in the United Kingdom, cooling is likely to remain a far

TABLE 3. The spread in values of projections from the four calibration methods, as presented in Fig. 7. Column 1 is the maximum spread in a single ensemble member across calibration methods (calibration uncertainty). Column 2 is the maximum spread in a single method across all models (model uncertainty).

Case	Maximum single-member spread	Maximum single-method spread
HDD, national	484	519
HDD, Glasgow	626	595
CDD, national	57	59
CDD, London	137	131

smaller concern than heating. However, the analysis in this paper does not account for trends or interannual variability. Since it is likely that individual hot years may result in high values of CDD even in a relatively cool climate, consideration of CDD in individual years is an important matter for future work. Moreover, temperature distributions in summer (and in some models, in winter) deviate from Gaussian, such that mapping the mean and variance alone as done in this paper will not fully capture changes in the tails and therefore in degree-days, so focusing on higher-order statistics or quantile mappings would be important for future work.

Note that the above discussion focuses on relative differences between methods and relative change in degree-days. In absolute terms, the uncertainties in HDD are much larger than those in CDD (Table 3), but these uncertainties are small relative to the total number of HDD. For some users, absolute values may be more relevant.

Outside the United Kingdom, in hot climates ($\mu > 22.0^\circ$), the findings in this paper regarding linearity may become critical for CDD projections. For example, simple index correction could result in unrealistic projections of future increases in degree-days (by analogy with case 2; section 3b). The specific analytical model derived here would only apply if the Gaussian assumption was valid.

A possible limitation of this analysis is the relevance of the thresholds used and their interaction with other socioeconomic factors (Isaac and Van Vuuren 2009). For example, cooling equipment is not currently widespread in the United Kingdom, so even hot days would not see energy use for cooling. If it were available, buildings might be cooled at less than 22°C ; Brown et al. (2016) recently found that in the United States, lower thresholds are appropriate for cooler regions and vice versa. Such findings also imply that the appropriate thresholds to use in degree-day analysis are likely to change over time, as people and societies adapt to the increasing temperatures resulting from climate change. Our analysis could easily be applied to different thresholds. The analytical expressions of Eqs. (4) and (5) would also enable simple allowance to be made for processes not captured in the model. For example, if the local temperature effect of an urban heat island was observed in reality but not simulated in a climate model, the effect on CDD could be easily explored by adjusting the mean in the expression.

This framework can be extended to other indices. Growing degree-days are an above-threshold degree-day with threshold 5.5°C (e.g., Harding et al. 2015). Given the 3-sigma rule of thumb for linearity discussed in this paper, and taking an approximate within-month standard deviation of 3°C , growing degree-days would be expected to be nonlinear in the mean μ and dependent upon standard deviation for

$\mu < (5.5 + 3 \times 3)^\circ\text{C} = 14.5^\circ\text{C}$. Therefore, the nonlinearity of growing degree-days is broadly relevant in the United Kingdom, particularly in the north. This is an important caveat to studies such as Harding et al. (2015), which calculate changes in the index alone. For a threshold exceedance index with threshold T_A , the expected value of the index from an equivalent analytical derivation is $(1/2) \{1 + \text{erf}[(T_A - \mu)/\sigma\sqrt{2}]\}$. Analysis of the limits of this function can reveal the cases in which threshold exceedance are nonlinear in μ and σ . We conclude then that this study provides valuable insight into the types of indices, across a range of impact sectors, that are particularly vulnerable to uncertainty arising from bias correction methodology.

Acknowledgments. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>; Haylock et al. 2008). E-OBS is available for download by registration (at www.ecad.eu/download/ensembles/ensembles.php). Caroline Holmes and Adam Butler were funded by Scottish Government through ClimateXChange (www.climateexchange.org.uk). Simon Tett is funded by the University of Edinburgh. The authors thank Kate Beckmann, formerly of Heriot-Watt University, for early discussions on this work; and Ed Hawkins of the University of Reading whose blog article “How not to use daily CMIP5 data for impact studies” (www.climate-lab-book.ac.uk, 2015) inspired the thinking behind looking carefully at threshold-based indices. The authors also thank three anonymous reviewers for their constructive comments, which greatly improved the manuscript, in particular regarding adding detail to the derivation in the appendix.

APPENDIX

Derivation of Analytical Expression

The derivation is given for the HDD case. Starting with Eq. (3),

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} (T_H - x)H(T_H - x)f_X(x)dx, \quad (\text{A1})\end{aligned}$$

and substituting the Gaussian distribution function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-[(x-\mu)^2/2\sigma^2]} \quad (\text{A2})$$

gives

$$\begin{aligned}
\mathbb{E}(Y) &= \int_{-\infty}^{\infty} (T_H - x) H(T_H - x) \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx \\
&= \int_{-\infty}^{T_H} (T_H - x) \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx \\
&= T_H \int_{-\infty}^{T_H} \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx - \int_{-\infty}^{T_H} x \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx \\
&= (T_H - \mu) \int_{-\infty}^{T_H} \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx - \int_{-\infty}^{T_H} (x - \mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx, \quad (\text{A3})
\end{aligned}$$

where the final expansion is made for convenience in what follows.

Using the substitution

$$t = \frac{(x - \mu)}{\sigma\sqrt{2}} \quad (\text{A4})$$

this becomes

$$\begin{aligned}
\mathbb{E}(Y) &= (T_H - \mu) \int_{-\infty}^{\frac{T_H - \mu}{\sigma\sqrt{2}}} \frac{1}{\sqrt{\pi}} e^{-t^2} dt - \int_{-\infty}^{\frac{T_H - \mu}{\sigma\sqrt{2}}} \frac{\sigma\sqrt{2}}{\sqrt{\pi}} t e^{-t^2} dt \\
&= \frac{(T_H - \mu)}{2} \left(1 + \operatorname{erf} \left[\frac{T_H - \mu}{\sigma\sqrt{2}} \right] \right) + \frac{\sigma}{\sqrt{2\pi}} e^{-[(T_H - \mu)^2/2\sigma^2]}, \quad (\text{A5})
\end{aligned}$$

using the results

$$1 + \operatorname{erf}(x) = \int_{-\infty}^x \frac{2}{\sqrt{\pi}} e^{-y^2} dy \quad (\text{A6})$$

$$\int_{-\infty}^x y e^{-y^2} dy = -\frac{e^{-y^2}}{2}. \quad (\text{A7})$$

For CDD, the integral to be evaluated becomes

$$\mathbb{E}(Y) = \int_{T_c}^{\infty} (x - T_c) \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx, \quad (\text{A8})$$

which through the same arguments gives the expression

$$\begin{aligned}
\mathbb{E}(Y) &= \frac{(\mu - T_c)}{2} \left(1 + \operatorname{erf} \left[\frac{\mu - T_c}{\sigma\sqrt{2}} \right] \right) \\
&\quad + \frac{\sigma}{\sqrt{2\pi}} e^{-[(\mu - T_c)^2/2\sigma^2]}. \quad (\text{A9})
\end{aligned}$$

REFERENCES

- Brown, M. A., M. Cox, B. Staver, and P. Baer, 2016: Modeling climate-driven changes in U.S. buildings energy demand. *Climatic Change*, **134**, 29–44, doi:10.1007/s10584-015-1527-7.
- Brown, S., P. Boorman, and J. Murphy, 2010: Interpretation and use of future snow projections from the 11-member Met Office Regional Climate Model ensemble. Met Office UKCP09 Tech. Note, 25 pp., <http://ukclimateprojections.metoffice.gov.uk/media.jsp?mediaid=87949&filetype=pdf>.
- Chow, D. H., and G. J. Levermore, 2010: The effects of future climate change on heating and cooling demands in office buildings in the UK. *Build. Serv. Eng. Res. Tech.*, **31**, 307–323, doi:10.1177/0143624410371284.
- Christenson, M., H. Manz, and D. Gyalistras, 2006: Climate warming impact on degree-days and building energy demand in Switzerland. *Energy Convers. Manage.*, **47**, 671–686, doi:10.1016/j.enconman.2005.06.009.
- Collins, M., B. B. Booth, B. Bhaskaran, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb, 2011: Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles. *Climate Dyn.*, **36**, 1737–1766, doi:10.1007/s00382-010-0808-0.
- Day, A., P. Jones, and G. Maidment, 2009: Forecasting future cooling demand in London. *Energy Build.*, **41**, 942–948, doi:10.1016/j.enbuild.2009.04.001.
- Erhardt, R., 2015: Mid-twenty-first-century projected trends in North American heating and cooling degree days. *Environmetrics*, **26**, 133–144, doi:10.1002/env.2318.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168, doi:10.1007/s003820050010.
- Hadley Centre for Climate Prediction and Research, 2008: UKCP09: Met Office Hadley Centre Regional Climate Model (HadRM3-PPE) data. NCAS British Atmospheric Data Centre, accessed 1 August 2013, <http://catalogue.ceda.ac.uk/uuid/f9c7d70b0e6d4c31c067e75ab67cb3cc>.
- Hanlon, H., G. Hegerl, S. Tett, and D. Smith, 2014: Near-term prediction of impact-relevant extreme temperature indices. *Climatic Change*, **132**, 61–76, doi:10.1007/s10584-014-1191-3.
- Harding, A., M. Rivington, M. Mineter, and S. Tett, 2015: Agro-meteorological indices and climate model uncertainty over the UK. *Climatic Change*, **128**, 113–126, doi:10.1007/s10584-014-1296-8.
- Hawkins, E., 2015: How not to use daily CMIP5 data for impact studies. Climate Lab Book: Open Climate Science, accessed 1 September 2015, <https://www.climate-lab-book.ac.uk/2015/how-not-to-use-daily-cmip5-data-for-impact-studies/>.
- , T. E. Fricker, A. J. Challinor, C. A. Ferro, C. K. Ho, and T. M. Osborne, 2013: Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. *Global Change Biol.*, **19**, 937–947, doi:10.1111/gcb.12069.
- Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data

- set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, doi:[10.1029/2008JD010201](https://doi.org/10.1029/2008JD010201).
- Ho, C. K., D. B. Stephenson, M. Collins, C. A. T. Ferro, and S. J. Brown, 2012: Calibration strategies: A source of additional uncertainty in climate change projections. *Bull. Amer. Meteor. Soc.*, **93**, 21–26, doi:[10.1175/2011BAMS3110.1](https://doi.org/10.1175/2011BAMS3110.1).
- Hofstra, N., M. Haylock, M. New, and P. D. Jones, 2009: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *J. Geophys. Res.*, **114**, D21101, doi:[10.1029/2009JD011799](https://doi.org/10.1029/2009JD011799).
- Holmes, C., 2017: Heating and cooling degree day projections for the UK: Version 2. Edinburgh DataShare, accessed 4 September 2017, doi:[10.7488/ds/2130](https://doi.org/10.7488/ds/2130).
- Isaac, M., and D. P. Van Vuuren, 2009: Modeling global residential sector energy demand for heating and air conditioning in the context of climate change. *Energy Policy*, **37**, 507–521, doi:[10.1016/j.enpol.2008.09.051](https://doi.org/10.1016/j.enpol.2008.09.051).
- Jenkins, G., M. Perry, and J. Prior, 2008: The climate of the UK and recent trends. Met Office Hadley Centre Tech. Rep., 122 pp., http://www.ukcip.org.uk/wp-content/PDFs/UKCP09_Trends.pdf.
- Kendon, M., M. McCarthy, S. Jevrejeva, and T. Legg, 2016: State of the UK Climate 2015. Met Office Hadley Centre Tech. Rep., 66 pp., https://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/r/4/state_of_the_uk_climate_2015.pdf.
- Lambert, F. H., G. R. Harris, M. Collins, J. M. Murphy, D. M. H. Sexton, and B. B. Booth, 2013: Interactions between perturbations to different Earth system components simulated by a fully-coupled climate model. *Climate Dyn.*, **41**, 3055–3072, doi:[10.1007/s00382-012-1618-3](https://doi.org/10.1007/s00382-012-1618-3).
- Lemonsu, A., R. Koukoku-Arnaud, J. Desplat, J.-L. Salagnac, and V. Masson, 2013: Evolution of the Parisian urban climate under a global changing climate. *Climatic Change*, **116**, 679–692, doi:[10.1007/s10584-012-0521-6](https://doi.org/10.1007/s10584-012-0521-6).
- Li, D. H., L. Yang, and J. C. Lam, 2012: Impact of climate change on energy use in the built environment in different climate zones—A review. *Energy*, **42**, 103–112, doi:[10.1016/j.energy.2012.03.044](https://doi.org/10.1016/j.energy.2012.03.044).
- Maraun, D., 2016: Bias correcting climate change simulations—A critical review. *Curr. Climate Change Rep.*, **2**, 211–220, doi:[10.1007/s40641-016-0050-x](https://doi.org/10.1007/s40641-016-0050-x).
- Murphy, J., and Coauthors, 2010: UK climate projections science report: Climate change projections. Met Office Hadley Centre Tech. Rep., 192 pp., <http://ukclimateprojections.metoffice.gov.uk/media.jsp?mediaid=87894&filetype=pdf>.
- Nakicenovic, N., and R. Swart, Eds., 2000: *Special Report on Emissions Scenarios*. Cambridge University Press, 612 pp.
- Perry, M., and D. Hollis, 2005: The development of a new set of long-term climate averages for the UK. *Int. J. Climatol.*, **25**, 1023–1039, doi:[10.1002/joc.1160](https://doi.org/10.1002/joc.1160).
- Ruiz-Ramos, M., A. Rodríguez, A. Dosio, C. M. Goodess, C. Harpham, M. I. Mínguez, and E. Sánchez, 2016: Comparing correction methods of RCM outputs for improving crop impact projections in the Iberian Peninsula for 21st century. *Climatic Change*, **134**, 283–297, doi:[10.1007/s10584-015-1518-8](https://doi.org/10.1007/s10584-015-1518-8).
- Sanderson, M. G., A. J. Wiltshire, and R. A. Betts, 2012: Projected changes in water availability in the United Kingdom. *Water Resour. Res.*, **48**, W08512, doi:[10.1029/2012WR011881](https://doi.org/10.1029/2012WR011881).
- Sniffer, 2014: Scotland's climate trends handbook: Sniffer project CC13. Scotland's Environment Tech. Rep., accessed 23 September 2016, http://www.environment.scotland.gov.uk/climate_trends_handbook.
- Taylor, J. W., and R. Buizza, 2003: Using weather ensemble predictions in electricity demand forecasting. *Int. J. Forecast.*, **19**, 57–70, doi:[10.1016/S0169-2070\(01\)00123-6](https://doi.org/10.1016/S0169-2070(01)00123-6).
- Trewin, B., 2007: *The Role of Climatological Normals in a Changing Climate*. World Meteorological Organization Rep. WCDMP 61, WMO-TD 1377, 46 pp., https://www.wmo.int/datastat/documents/WCDMPNo61_1.pdf.
- Weisheimer, A., and T. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, doi:[10.1098/rsif.2013.1162](https://doi.org/10.1098/rsif.2013.1162).
- Zubler, E. M., S. C. Scherrer, M. Croci-Maspoli, M. A. Liniger, and C. Appenzeller, 2014: Key climate indices in Switzerland; expected changes in a future climate. *Climatic Change*, **123**, 255–271, doi:[10.1007/s10584-013-1041-8](https://doi.org/10.1007/s10584-013-1041-8).